

# Comparison analysis of Bangla news articles classification using support vector machine and logistic regression

Md Gulzar Hussain<sup>1,2</sup>, Babe Sultana<sup>1</sup>, Mahmuda Rahman<sup>1</sup>, Md Rashidul Hasan<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Faculty of Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh

<sup>2</sup>Department of Computer Application Technology, School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu, China

---

## Article Info

### Article history:

Received Feb 21, 2022

Revised Oct 29, 2022

Accepted Nov 12, 2022

### Keywords:

Bangla news

Bangla text classification

Logistic regression

Natural language processing

News classification

Support vector machine

## ABSTRACT

In the information age, Bangla news articles on the internet are fast-growing. For organizing, every news site has a particular structure and categorization. News article classification is a method to determine a document's classification based on various predefined categories. This research discusses the classification of Bangla news articles on the online platform and tries to make constructive comparison using several classification algorithms. For Bangla news articles classification, term frequency-inverse document frequency (TF-IDF) weighting and count vectorizer have been used as a feature extraction process, and two common classifiers named support vector machine (SVM) and logistic regression (LR) employed for classifying the documents. It is clear that the accuracy of the experimental results by applying SVM is 84.0% and LR is 81.0% for twelve categories of news articles. In this research work, when we have made comparison two renowned classification algorithms applied on the Bangla news articles, LR was outperformed by SVM.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



---

## Corresponding Author:

Md Gulzar Hussain

Department of Computer Science and Engineering, Faculty of Science and Engineering

Green University of Bangladesh, Dhaka, Bangladesh

Email: gulzar.ace@gmail.com

---

## 1. INTRODUCTION

Text data is the most comprehensive source of information but due to its unstructured nature, it is difficult and time consuming to draw insights from it. Advancement in machine learning (ML) and natural language processing (NLP) are making it easier to analyze text data through text classification. Text classification techniques are used to organize, structure and categorize text data for analysis of sentiment, topic labelling, spam detection, intent detection and so on [1]. This study applies text classification techniques to Bangla news articles as the news articles are the most common form of text online. Many studies have been conducted on text classification, and different classification techniques such as rules-based method, decision trees, k-nearest neighbors (KNN), naive Bayes, logistic regression, support vector machines (SVM), neural networks (NN) and so on have been developed [2]. Studies in the literature of various research papers have shown that researchers have focused on classifying texts in different languages, such as English and Arabic and so on [3], [4] however, the amount of analysis on the Bengali language is much less. In [5] have used machine learning algorithms to categorize Bangla newspaper articles into five distinct groups. They used logistic regression, SVM, and multi-layer neural network to do so. Multi-layer dense neural network approach has a

accuracy of 95.50% which is higher compared to other models. There is a representation [6] of the behavior of least-squares SVMs, twin SVMs, and least-squares twin SVM (LS-TWSVM) classifiers on the news data is shown to handle multi-category data. And their performance evaluation showed that LS-TWSVM is the best of all three with 92.96% accuracy. Maisha *et al.* [7] perform sentiment analysis on Bangla news using "pipeline" class along with six state-of-the-art supervised ML algorithms which includes decision tree (DT), multinomial naive Bayes (MNB), k-nearest neighbor (KNN), logistic regression (LR), random forest (RF) and lagrangian support vector machine (LSVM). Random forest algorithm out stands all other algorithms securing 98% accuracy in percentage split method. This methodology categorizes the content either as positive or negative. Rabbinnov and Kobilov [8] used SVM, decision tree classifier, random forest, LR, and MNB among six other machine-learning algorithms to conduct multi-class text categorization of internet Uzbek news articles. For SVM, they used radial basis function (RBF SVM) which gave the best accuracy (86.88%) performance of other classifiers. Several supervised machine learning as well as deep learning algorithms for categorizing Bengali news documents are discussed in [9]. A never method for classifying Bangla textual content has been developed by [10]. The deep learning recurrent neural networks (RNN) - based attention layer and the RNN with BiLSTM achieved accuracy rates of 97.72% and 86.56%, respectively. In [11] proposed a model structure named the DCLSTM-MLP model for the categorization of news text documents, an idea of a customized algorithm that combines deep learning algorithms such as long term short memory (LSTM), convolutional neural network (CNN), and multi-layer perceptron (MLP). By applying this model structure they have tried to solve the problems of textual length, the complexity of extracting features from news content and categorizing news text effectively and achieved 94.82% accuracy. However, the main issue of their paper is that the number of samples is small, and the distribution of different types of news is unequal, resulting in the model's narrow effectiveness.

Kowsher *et al.* of [12] used several word embedding methods to incorporate the text from Bangla newspaper data, as well as machine learning algorithms to categorize the incorporated text. A deep recurrent neural Network was used to provide a new method of evaluating Bangla news articles in [13]. The deep recurrent neural network featuring Bi-LSTM obtained 98.33% in Bengali text categorization, which is greater than previous well-known classification techniques. For classifying text or documents, many of supervised techniques are used such as KNN, naive Bayes (NB), DT, n-grams, neural networks (NNet). But according to previous research literature reviews SVM [14] is most frequently used classifier algorithm. In [15] proposed a news article classification model framework based on deep hybrid learning and compared it to traditional text classification to demonstrate the superiority of network news text classification, and it outperforms the standardized method for classifying news texts in terms of overall performance. In [16] showed a comparative analysis among DT, KNN, NB, and Rocchio's algorithm where their studies said that SVM outperforms better than all other classifiers. Aside from the English document, there has also been extensive research on other languages. In [17] focused on the classification of self-created Indonesian news corpus, which includes four separate categories and 472 Indonesian newspaper articles from a variety of sources. They produced five models with ten epoch each using 377 data for training and 95 testing data, with CNN having the highest accuracy rate at about 90.74 percent categorizing Indonesian news data. In [18] conducted on an Indonesian news corpus found that the combination of TF-IDF and MNB beats other classification models such as multivariate Bernoulli naive Bayes (BNB) and SVM with an accuracy of 85%. In terms of Arabic text classification, in [19] Arabic medical text documents are classified and authors used an rule-based classifier for classification of Arabic text and having an accuracy of 90.6%. They used three classification algorithm: majority voting, ordered decision list, and K-NN which are used to validate the model.

Among all the couple works covered in the Bangla, Chy *et al.* [20] worked on Bangla language news classification and they used naive Bayes classifier. But the problem is that in the naive Bayes model if the testing data set contains a categorical variable of a category that was not part of the training data set, it will give it zero probability and be unable to make any predictions. Nahar *et al.* [21] demonstrated a comparative analysis utilizing naive Bayes classifier, SVM, and neural networks to filter Bangla sports and political news only on online networks from text data. Mandal and Sen [22] showed the comparison performance evaluation of the 4 supervised learning techniques: decision tree, naive Bayes, k-nearest neighbor, and SVM for the Bengali text categorization. They used TF-IDF to comprise a feature vector using 1000 web documents which contains 22,218 words. Close to our research work, by using TF-IDF as a feature selection approach and SVM as a classifier, they attained a classification accuracy of 92.57 percent for 12 categories of Bangla text files. They used 3191 text samples per category where we have tried to use 11770 articles of top 12 categories (showed

in Table 1) from 20 different categories. And best of our knowledge they used only SVM where we have used SVM and logistic regression (LR) and LR is primarily used to classify observations into a discrete number of categories with rapid classification of unknown records. In this research paper, we have tried to show a comparative performance analysis using SVM and LR which can bring righteous path for future researchers who are willing to work with Bangla language in this era.

In this study, LR and SVM are used to classify the Bangla news articles as the study [23] shows that the logistic regression outperforms other techniques such as random forest and k nearest neighbours algorithm. It classifies the Bangla news articles and label them to a certain topic based on their contents. It extracts features from the corpus using the TF-IDF feature vector since it is enabled to the extraction of relevant features as well as removing common words.

## 2. METHOD

The aim of news classification seems to allocate categories as per the content of a news article. Pre-processing of Bangla news articles and feature set extraction are also needed prior to training and model construction for document classification, like English text classification. The overall Bangla news article classification process have been used in this experiment as illustrated in Figure 1.

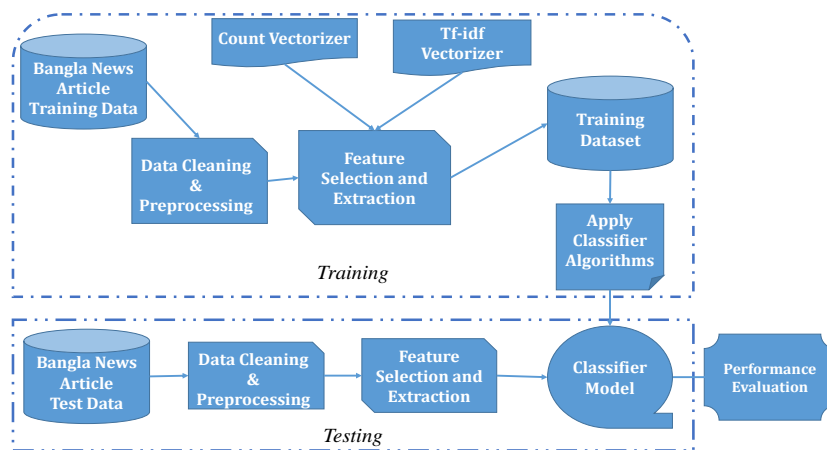


Figure 1. System architecture of the proposed work

### 2.1. Data collection

Despite the limited amount of the data, Hossain *et al.* [24] discovered that text-graph convolutional neural (GCN) accomplished better than GRU-LSTM, BiLSTM, Char-CNN, LSTM, and bidirectional encoder representations from transformers (BERT) in classifying online Bangla news. As so far, there is a scarcity of standard data set in the Bangla language, so a data set have been prepared by scrapping the news articles from various electronic news sites such as '<https://www.prothomalo.com/>'. At the time of scrapping, the articles were labeled with their categories. To assess the recommendation results, we have collected around 12.5 K labeled news articles consisting of 20 categories. Among them top 12 categories are considered for this research work which has mentioned in Table 1.

Table 1. Category-wise count of the Bangla news articles

Serial	Category	Articles	Serial	Category	Articles
1	Bangladesh news	5029	7	Feedback news	578
2	Sports news	1387	8	Other news	447
3	North America news	809	9	Citizen news	414
4	Entertainment news	750	10	News of distance migrants	393
5	International news	746	11	Lifestyle news	264
6	Economic news	711	12	Science and technology news	242

## 2.2. Data preprocessing

Preprocessing is performed to minimize the noise in the text, which helps to increase the classifier's efficacy. The preprocessing steps clear the text data and preparing it for the artificial learning model. Then the tokenization is performed on text documents and break down the sequence of characters or words to get the features. Since text documents contain sentences, like a sequence of character or words, to get the features, it should be broken into tokens. Tokenization is the process to breakdown the text into sentences and then words. Tokenization is the process by which the text is divided into phrases and then words. After tokenization, different symbols such as !, x, i, j, \$, %, and numbers, which are not very important for classification, are excluded. Stop-words in Bangla are also eliminated at the same period. In [25] is a list of stop-words used in this research.

## 2.3. Feature extraction

The general method of conversion of a collection of text documents to numerical feature vectors is vectorization. Count vectorizer transforms a text document array into a token counting matrix: contains token counts occurrences in each document. A sparse representation of the numbers is generated by this implementation. On the other hand, the purpose of using TF-IDF instead of the raw frequencies of a token's occurrence in a given document is to scale down the influence of tokens that occur in a given corpus very regularly. And thus empirically less informative characteristics that occur in a small fraction of the training corpus are removed. It is a method of data retrieval that weights the frequency of words (TF) and the inversed document frequency (IDF). Every word has it's own TF and IDF ratings, respectively. The TF-IDF weight is the product of a term's TF and IDF scores. The rarer the word and vice versa, the higher the TF-IDF score.

## 2.4. Classifiers

In this study, two common classifiers: the SVM classifier with a linear kernel and the logistic regression have been used. In text classification, SVM has been used effectively. And logistic regression is a statistical method of data analysis in which one or more variables are used to determine the result.

### 2.4.1. SVM

In essence, SVM is a supervised machine learning approach known as a binary classifier. In [25] uses hyper-plane to classify data into two types. Support vectors are closer to the-hyper-plane data points that influence the position and orientation of the hyper-plane. The support vector is used to optimize the perimeter of the classifier. By eliminating the support vectors, the hyper-plane's position will change. SVM recognizes the following sign function of equation [26] mathematically,

$$F(x) = \sin(wx + b) \quad (1)$$

where  $w$  is an  $n$  weighted vector in  $R^n$ . By dividing the space  $R^n$  into two half-spaces with the maximum margin, SVM finds the hyper-plane in (2).

$$y = wx + b \quad (2)$$

Generally, SVM is a binary classifier, but the strategies like one-to-one and one-vs-rest can be used to expand into a multi-class classifier. In SVM, linear and radial basis function kernels are used to make decisions. Since most texts are linearly separable, the linear kernel is preferred for text classification.

### 2.4.2. Logistic regression

Logistic regression uses a logistic function to estimate probabilities for the relationship between one or more independent variables and the dependent variable of the categories. The logistic function equation is also known as the logistic curve which is a common "S" shaped curve described by the (3). The sigmoid curve is another name for the logistic curve.

$$F(x) = \frac{L}{1 + e^{-k(v-v_0)}} \quad (3)$$

Where,  $L$  is the curve's highest value,  $e$  the base of a natural logarithm (or euler's number),  $k$  is the logistic growth rate or curve steepness,  $v_0$  is the sigmoid midpoint's  $x$ -value.

### 3. RESULTS AND DISCUSSION

The performance of SVM classifiers and logistic regression on our data-set is discussed in this section. A little analysis on the data-set is provided here. Later, the performance analysis of SVM and LR on the Bangla text data and comparison with some similar works are demonstrated in Table 4.

#### 3.1. Dataset analysis

Among 12.5 thousand of articles with 20 categories we used 11770 articles of top 12 categories in this research work. Bar-chart for number of news articles of each categories are given in Figure 2. In Figure 3 generated word cloud based on TF-IDF and count vectorizer from total articles are given. This is generated before running the preprocessing on the data.

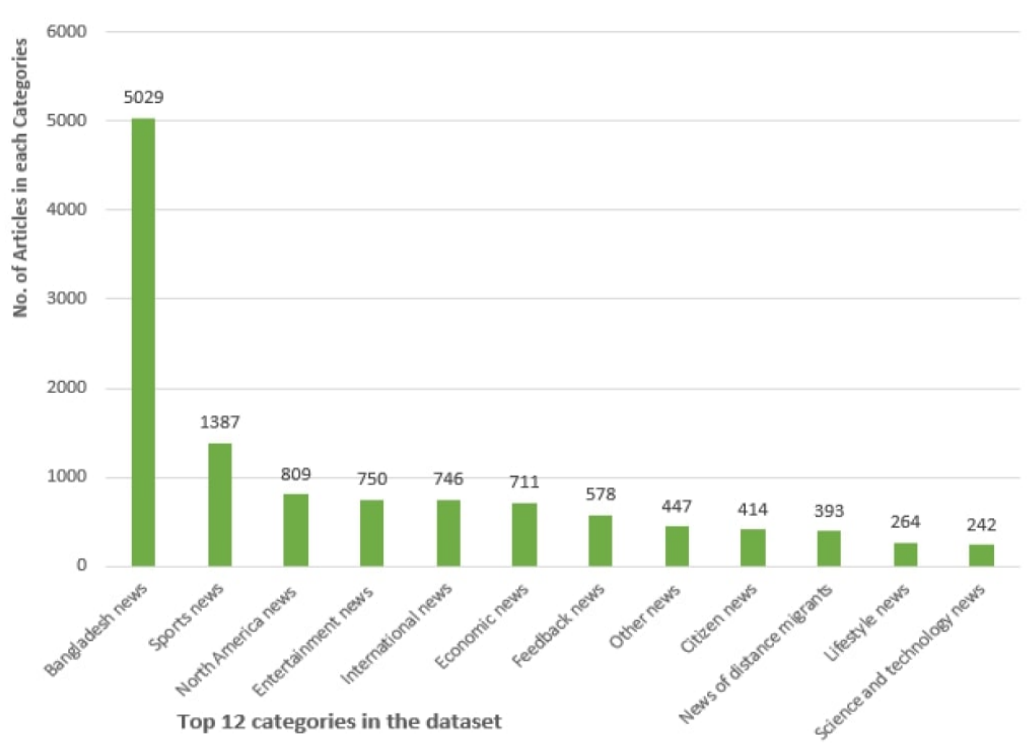


Figure 2. Numbers of the Bangla news articles for each category



Figure 3. Word cloud for the articles used based on TF-IDF (left) and count vectorizer (right)

We can see many unnecessary words are available in the clouds and the cloud-based on TF-IDF is different than the cloud-based on count vectorizer value. But some words are having around the same importance for both TF-IDF and count vectorizer value. We have split the data into two parts, for training the model, 80% of the data is used and the remaining 20% is used to test the performance.

### 3.2. Performance measures

In-text categorization, a variety of evaluation metrics are employed. The precision, recall, and F-measure are among the most widely used performance measures considered in our experiments. Each category's precision, recall, and F-measure experiment are measured for SVM and LR. The F-measure is averaged to assess performance throughout categories. Micro average and macro average are the two types of average value and the macro average is used. Table 2 show how two classifiers performed on our data-set corpus in terms of precision, recall, and F-measure. On the other hand Table 3 shows comparison of accuracy, average precision, average recall and average F1-score result for SVM and LR classifier.

Table 2. SVM and LR classifier results of 12 categories

Categ.	Precision		Recall		F1-Score		Categ.	Precision		Recall		F1-Score	
	SVM	LR	SVM	LR	SVM	LR		SVM	LR	SVM	LR	SVM	LR
0	0.89	0.84	0.96	0.96	0.92	0.90	6	0.69	0.64	0.74	0.69	0.71	0.66
1	0.94	0.91	0.98	0.98	0.96	0.95	7	0.77	0.78	0.78	0.76	0.77	0.77
2	0.74	0.78	0.72	0.68	0.73	0.73	8	0.51	0.50	0.21	0.13	0.29	0.21
3	0.88	0.84	0.87	0.83	0.88	0.84	9	0.59	0.60	0.49	0.41	0.53	0.49
4	0.70	0.65	0.68	0.61	0.69	0.63	10	0.88	0.93	0.61	0.49	0.72	0.64
5	0.81	0.80	0.78	0.73	0.79	0.76	11	0.69	0.68	0.57	0.43	0.62	0.53

It is observed that SVM with linear kernel achieves accuracy of 0.84 and LR achieves 0.81 from Table 3 and SVM outperforms LR in case of average precision, recall and F1-score. Though from Table 2 it is found that for some classes or categories LR outperforms SVM in case of individual precision, recall and F1-score. Figure 4 shows the confusion matrix for SVM classifier with the linear kernel and logistic regression respectively. Table 4 shows the comparison between the recent works using SVM classifier findings with the results of this research finding. The comparison shows that our research work utilizing the combination of TF-IDF and count vectorizer features in the SVM classifier achieves better accuracy than other recent works which also used SVM in their research works.

Table 3. Accuracy, average precision, average recall and average F1-score result of SVM and LR classifier

Accuracy		Precision (Avg)		Recall (Avg)		F1-score (Avg)	
SVM	LR	SVM	LR	SVM	LR	SVM	LR
0.84	0.81	0.76	0.75	0.70	0.64	0.72	0.68

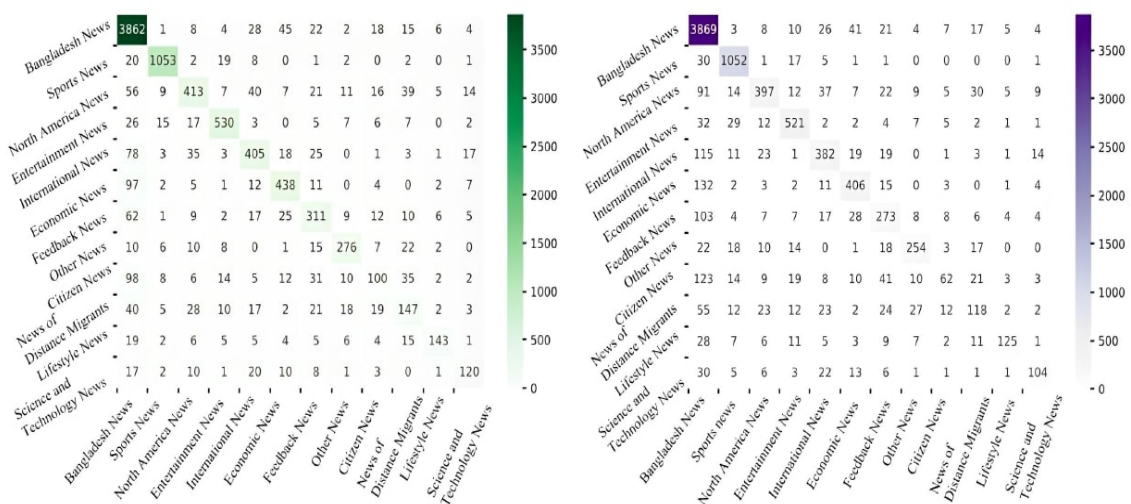


Figure 4. Confusion-matrix for SVM (left) and LR (right)

Table 4. Comparison between this experiment and recent similar researches using SVM

Experiments	Feature used	Accuracy
This experiment	TF-IDF + Countvectorizer	0.84
Islam <i>et al.</i> [27]	Trigram TF-IDF	0.80
Rahman <i>et al.</i> [28]	TF-IDF	0.82
Yeasmin <i>et al.</i> [5]	TF-IDF	0.83

#### 4. CONCLUSION

In the field of information systems, text categorization is a contentious issue. In this paper, we have used SVM and LR classifiers on own developed corpus and their performance is measured. The proposed methodology in this study advocates the assumption that the Bangla language can indeed be rightly classified using SVM and LR with limited resources. The outcome of the suggested approach is promising. Still, more precision could have attained. Good outcomes could also have obtained if we had been able to deal with all of the news categories and utilize multiple categorization methods to come with a constructive opinion. In the next, we'd like to add more categories and compare the results using different classification algorithms.

#### ACKNOWLEDGEMENT

This work was supported in part by the Center for Research, Innovation, and Transformation of Green University of Bangladesh.




#### REFERENCES

- [1] X. Zhou *et al.*, "A survey on text classification and its applications," *Web Intelligence*, vol. 18, no. 3, pp. 205-216, 2020, doi: 10.3233/WEB-200442.
- [2] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, 2019, doi: 10.3390/info10040150.
- [3] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, "Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya," *Information*, vol. 12, no. 2, 2021, doi: 10.3390/info12020052.
- [4] K. T. Nwet, "Machine learning algorithms for myanmar news classification," *International Journal on Natural Language Computing*, vol. 8, no. 4, pp. 17–24, 2019, doi: 10.5121/ijnlc.2019.8402.
- [5] S. Yeasmin, R. Kuri, A. R. M. M. H. Rana, A. Uddin, A. Q. M. S. U. Pathan, and H. Riaz, "Multi- category bangla news classification using machine learning classifiers and multi-layer dense neural network," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021, doi: 10.14569/IJACSA.2021.0120588.
- [6] P. Saigal and V. Khanna, "Multi-category news classification using support vector machine based classifiers," *SN Applied Sciences*, vol. 2, 2020, doi: 10.1007/s42452-020-2266-6.
- [7] S. J. Maisha, N. Nafisa, and A. K. M. Masum, "Supervised machine learning algorithms for sentiment analysis of Bangla newspaper," *International Journal of Innovative Computing*, vol. 11, no. 2, pp. 15–23, 2021, doi: 10.11113/ijic.v11n2.321.
- [8] I. M. Rabbimov and S. S. Kobilov, "Multi-class text classification of uzbek news articles using machine learning," in *Journal of Physics: Conference Series*, 2020, doi: 10.1088/1742-6596/1546/1/012097.
- [9] M. R. Hossain, S. Sarkar, and M. Rahman, "Different machine learning based approaches of baseline and deep learning models for bengali news categorization," *International Journal of Computer Applications*, vol. 176, no. 18, pp. 10-16, doi: 10.5120/ijca2020920107.
- [10] M. Ahmed, P. Chakraborty, and T. Choudhury, "Bangla Document Categorization Using Deep RNN Model with Attention Mechanism," *Cyber Intelligence and Information Retrieval*, vol. 291, pp. 137–147, 2021.
- [11] M. Zhang, "Applications of deep learning in news text classification," *Scientific Programming*, vol. 2021, 2021, doi: 10.1155/2021/6095354.
- [12] M. Kowsher, A. Tahabilder, N. J. Prottasha, M. A-Rakib, M. M. Uddin, and P. Saha, "Bangla topic classification using supervised learning," *Computational Intelligence in Pattern Recognition*, vol. 1349, pp. 505–518, 2022.
- [13] S. Rahman and P. Chakraborty, "Bangla document classification using deep recurrent neural network with bilstm," in *PComputer Science*, 2021.
- [14] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.
- [15] N. Sun and C. Du, "News text classification method and simulation based on the hybrid deep learning model," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/8064579.
- [16] P. Y. Pawar and S. H. Gawande, "A comparative study on different types of approaches to text categorization," *International Journal of Machine Learning and Computing*, vol. 2, no. 4, pp. 423-426, 2012. [Online]. Available: <http://www.ijmlc.org/papers/158-C01020-R001.pdf>
- [17] M. A. Ramdhani, D. S. Maylawati, and T. Mantoro, "Indonesian news classification using convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 2, pp. 1000–1009, 2020, doi: 10.11591/ijeecs.v19.i2.pp1000-1009.
- [18] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, Rudy, "News article text classification in Indonesian language," *Procedia Computer Science*, vol. 116, pp. 137–143, 2017, doi: 10.1016/j.procs.2017.10.039.




- [19] Q. A. Al-Radaideh and S. S. Al-Khateeb, "An associative rule-based classifier for arabic medical text," *International Journal of Knowledge Engineering and Data Mining*, vol. 3, no. 3-4, pp. 255–273, 2015, doi: 10.1504/IJKEDM.2015.074071.
- [20] A. N. Chy, M. H. Seddiqui, and S. Das, "Bangla news classification using naive bayes classifier," *16th Int'l Conf. Computer and Information Technology*, 2014, pp. 366-371, doi: 10.1109/ICCITechn.2014.6997369.
- [21] L. Nahar, Z. Sultana, N. Jahan, and U. Jannat, "Filtering Bengali Political and Sports News of Social Media from Textual Information," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934605.
- [22] A. K. Mandal and R. Sen, "Supervised learning methods for bangla web document categorization," *International Journal of Artificial Intelligence Applications*, vol. 5, no. 5, 2014, doi: 10.5121/ijaia.2014.5508.
- [23] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Human Research*, vol. 5, pp. 1–16, 2020, doi: 10.1007/s41133-020-00032-0.
- [24] M. R. Hossain, M. M. Hoque, N. Siddique, and I. H. Sarker, "Bengali text document categorization based on very deep convolution neural network," *Expert Systems with Applications*, vol. 184, 2021, doi: 10.1016/j.eswa.2021.115394.
- [25] Genediazjr, "Stopwords-ISO/stopwords-BN: Bengali stopwords collection," in *Github*, Oct 2016. Online. [Available]: <https://github.com/stopwords-iso/stopwords-bn> (Accessed: December 20, 2022).
- [26] J. Kim, J. Yoo, H. Lim, H. Qiu, Z. Kozareva, and A. Galstyan, "Sentiment prediction using collaborative filtering," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, no. 1, pp. 685–688, 2013, doi: 10.1609/icwsm.v7i1.14461.
- [27] T. Islam, A. I. Prince, M. M. Z. Khan, M. I. Jabiullah, and M. T. Habib, "An in-depth exploration of Bangla blog post classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 742–749, 2021, doi: 10.11591/eei.v10i2.2873.
- [28] S. Rahman, S. K. Mithila, A. Akther and K. M. Alam, "An Empirical Study of Machine Learning-based Bangla News Classification Methods," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6, 2021, doi: 10.1109/ICCCNT51525.2021.9579655.

## BIOGRAPHIES OF AUTHORS






**Md Gulzar Hussain**    was born in Nimnagar, Dinajpur City. He received the Bachelor of Science degree (2018) in Computer Science and Engineering from the Green University of Bangladesh, Dhaka, Bangladesh. Since 2019, he has been a Lecturer with the Computer Science and Engineering Department, Green University of Bangladesh, Dhaka, Bangladesh. He is currently studying MSc in Computer Science and Technology in Changzhou University, Changzhou, Jiangsu, China. He is also affiliated with IEEE as student member and IAER as professional member. His research interests include Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, Text Mining, and Topic Modeling. He can be contacted at email: [gulzar.ace@gmail.com](mailto:gulzar.ace@gmail.com).






**Babe Sultana**    was born in, Cox's Bazar, Bangladesh, in 1994. She received her B.Sc. Degree in Computer Science and Engineering from Green University of Bangladesh (GUB) in 2018. At present she is working as a Lecturer, Dept. of CSE, Green University of Bangladesh. Also, her publication was about "Multi-mode Project Scheduling with Limited Resource and Budget Constraints" published in International Conference on Innovation in Engineering and Technology (ICIET) 27-28 December, 2018 and she got Best Paper Award and IEEE Best Paper Award on this conference. Her research interests include Theory of Optimization, Natural language Processing, Renewable and Sustainable Energy. She can be contacted at email: [babecse@gmail.com](mailto:babecse@gmail.com).



**Mahmuda Rahman**    received her Master's in CS from Jahangirnagar University and B.Sc. in CSE from Green University of Bangladesh. She has received Vice Chancellor's Gold Medal award for her outstanding academic performance in 4th convocation of Green University of Bangladesh. She has been working as a Lecturer, Dept. of CSE, Green University of Bangladesh since 2019. Her research interest includes Natural Language Processing, Medical image processing and Machine learning. She can be contacted at email: [aurthi018@gmail.com](mailto:aurthi018@gmail.com).



**Md Rashidul Hasan**    received the Bachelor of Science degree in Computer Science and Engineering from Green University of Bangladesh. He was born in Pabna, Bangladesh in 1994. His research interest include Machine Learning, Natural Language Processing, and Data Mining. He can be contacted at email: [rashidulhasan777@gmail.com](mailto:rashidulhasan777@gmail.com).